



TITLE:

A Study on Social Information Search and Analysis on the Web by Diversity Computation(Abstract_要旨)

AUTHOR(S):

Shoji, Yoshiyuki

CITATION:

Shoji, Yoshiyuki. A Study on Social Information Search and Analysis on the Web by Diversity Computation. 京都大学, 2015, 博士(情報学)

ISSUE DATE:

2015-03-23

URL:

<https://doi.org/10.14989/doctor.k19119>

RIGHT:

(続紙 1)

京都大学	博士（ 情報学 ）	氏名	莊司 慶行
論文題目	A Study on Social Information Search and Analysis on the Web by Diversity Computation (多様性計算に基づくウェブ上のソーシャル情報の検索と分析に関する研究)		
<p>(論文内容の要旨)</p> <p>本論文は、近年、ウェブ (Web) データの中で、ソーシャルメディア等の比較的品質の低いユーザ生成型投稿コンテンツの比率が増加していることを前提として、このようなデータから、重要性や信頼性の高いデータを検索するために必要となる検索・分析方式を提案している。学位申請者は、従来のウェブのハイパーリンクのみならず、コミュニケーションサービス上での発言とウェブサイト間の参照構造、レビューとアイテム間の評価関係など、ソーシャルメディアとウェブを横断する「拡張されたリンク」に着目し、このようなリンクの設定理由、リンクの作成者プロフィール、リンク構造で結合されたコンテンツ等の分析を行うことで、より信頼性の高いコンテンツを検索する手法を提案している。リンクに関連するコンテンツ分析においては、情報の信頼性を担保するために、これらのコンテンツの「多様性」が重要であると考え、そのために必要な「多様性計算」のための手法を提案し、提案手法の有効性検証のための評価実験を行い、興味深い結果を得ている。</p> <p>具体的には、本論文では「リアクションをクエリとする情報検索」、「多様性計算に基づくオンラインレビューの信頼性推定」、「参照先・参照元の多様性に基づくHITSアルゴリズム」に関する研究を行い、提案手法・システムの評価を行っている。本論文は全7章から構成されている。その概要は以下の通りである。</p> <p>第1章は序論であり、本研究の背景、アプローチ、および、本論文の研究の全体の概要について述べている。</p> <p>第2章では、第3～6章で取り組んだ各研究課題の関連研究について述べ、本研究との位置づけを述べている。</p> <p>第3章では、ソーシャルメディア上の情報を用いることで従来のウェブ検索を改善する手法を提案している。提案手法では、ソーシャルメディアデータとウェブを横断的に利用することで、検索システムのクエリとして、読み手の印象を入力可能にしている。現実では「面白いSFを紹介して欲しい」など印象を伝えて検索する 경우가多いが、従来の検索エンジンでは「面白い」などの印象語は低精度の原因となる。そこでTwitter上でユーザがウェブページを紹介された際に「なるほど」などの発言（リアクション）をとることに注目し、これらを収集することでページが読者にどのような印象を与えるかを推定する手法について提案している。提案された手法をもとに実装された検索システムは、例えば、「わかりやすい C言語」というクエリに対して、「多数のユーザが、わかりやすいと類似した意味の発言をし、なおかつ、本文中にC言語という語を含む文書」を発見するものである。評価実験を通して、提案手法が従来の検索システムに比べて、クエリが「かわいい」、「便利だ」、「美味しそう」などの印象・評価に関わるリアクションを含んでいる場合に、より高精度に多くのウェブページを検索可能であることを実証している。同時に、この評価実験を通して、類似リアクションを取ったユーザの多数性を考慮するだけでは検索性能としては不十分であることも明らかにしており、多数性に加えて、多様性分析が必要となることが強く示唆されている。</p>			

第4章では、オンライン投稿型レビューサイトに投稿されたアイテムレビューについて、多様性計算に基づく信頼性推定手法について提案している。レビューの信頼性としてレビュー個人およびグループ（レビュー集合）について、アイテムをどれだけ正確に評価できるかを信頼性として定義し、分析を行っている。電子商取引が一般化し投稿レビューの重要性が増しているが、現在の多くのレビューサイトでは専門家と非専門家かを区別せず、評点の信頼性を考慮していない。より高品質なレビュー情報の生成・活用のために、レビュー個人の信頼性を推定するための仮説として「多数多様なアイテムをレビューしたレビューは信頼性が高い」、複数のレビューからなるグループの信頼性を推定する仮説として「多数多様なレビューからなるグループは信頼性が高い」を挙げ、実際の映画レビューデータをもとに批評家の評点と比較分析、仮説検証を行っている。実験結果として、レビュー個人の信頼性はレビューのそれまでのアイテムレビュー経験の量・多様さに大きく依存すること、グループの信頼性については、グループの人数がグループの信頼性と大きく関係すること、グループが十分に大きくない場合にエントロピーに基づく集団の多様性計算が信頼性推定に有効であるとの知見を得ている。

第5章では、従来のリンク解析アルゴリズム（HITSアルゴリズム）に、多様性の概念を導入することで、より一般性・信頼性の高いコンテンツを発見するアルゴリズム（参照先・参照元多様性に基づくHITSアルゴリズム）を提案している。HITSアルゴリズムはリンクの本数のみに基づきウェブページのスコアを計算するものであるが、たとえ同一リンク数であっても多様な内容のページ集合からリンクされている場合と、類似ページの集合からリンクされている場合ではページの質が異なる。そこでページのトピックから多様性を計算し、リンク先ページ内容が類似する場合にHub値が上がらず、リンク元ページ内容が類似する場合のAuthority値が上がらないよう、アルゴリズムの拡張を行っている。これにより、偏った読者層を持つページや視野の狭いリンク集のスコアを下げ、幅広い読者層に有用なページを上位にランク付け可能であることを評価実験で示している。

第6章では、本研究で得られた研究成果をまとめ、さらに今後の展開について述べている。

注) 論文内容の要旨と論文審査の結果の要旨は1頁を38字×36行で作成し、合わせて、3,000字を標準とすること。

論文内容の要旨を英語で記入する場合は、400～1,100 wordsで作成し
審査結果の要旨は日本語500～2,000字程度で作成すること。

(続紙 2)

(論文審査の結果の要旨)

本論文は、ソーシャルメディアとしての側面を有するようになった現代のウェブを高度に検索・分析する手段として、リンクの背景情報に基づくリンク解析手法の提案を行ったものである。さらに、リンクに関連するコンテンツ分析においては、情報の信頼性を担保するために、これらのコンテンツの「多様性」が重要であると考え、そのために必要な「多様性計算」手法を提案し、提案手法の有効性検証のための評価実験を行い、興味深く、学術上重要な結果を得ている。

具体的には、本論文の研究によって得られた成果は以下のように要約される。

1. 従来のウェブのハイパーリンクのみならず、コミュニケーションサービス上での発言とウェブサイト間の参照構造、レビューとアイテム間の評価関係など、ソーシャルメディアとウェブを横断する「拡張されたリンク」概念を導入し、ソーシャルメディアと従来のウェブ情報を横断的に利用する手法を提案した。
具体的には、オンラインコミュニケーションサービス上でページを紹介された際にとるリアクションを収集することで読み手が文書に対して抱く印象や評価語を入力可能な検索モデルについて提案し、従来の本文中に登場する語に基づく検索モデルと比較実験を行い、その優位性を明らかにした。
2. 投稿型レビューサイトにおけるレビューの信頼性を、レビュー個人とグループのレビュー履歴を用いた多様性計算を行うことにより推定する手法について提案し、実際の映画レビューサイトのデータを用いた実験を行い、レビュー個人においてレビューしたアイテムの量、多様性が信頼度に大きく寄与し、またグループにおいてグループの大きさとグループのメンバーのレビュー履歴の偏りの小ささが信頼度に大きく寄与するという知見を得た。
3. 従来リンクの本数のみによって計算されていたHITSアルゴリズムに対して多様性計算を組み込むことで、「一部の専門家だけでなく、初級者にとっても有用」など、一般性・信頼性の高い文書を検索する手法を提案し、評価実験を通して、従来のHITSアルゴリズムと比較して、検索結果上位においてより高精度に適合文書を発見可能であることを示した。

以上、本論文は、多様性計算に基づいてウェブ上のソーシャル情報の検索と分析する手法について研究を行ったものである。申請者は、本研究を通して、ソーシャルメディアと従来のウェブを横断的に用いる手法、社会科学の知見である多様性に基づく情報検索・分析手法を提案した。さらに、提案手法の有効性を示すために詳細な実験を行うなど、学術上寄与するところが少なくない。よって、本論文は、博士(情報学)の学位論文として価値あるものと認める。また、平成27年2月27日に論文内容とそれに関連した事項について試問を行った結果、合格と認めた。

注) 論文審査の結果の要旨の結句には、学位論文の審査についての認定を明記すること。
更に、試問の結果の要旨(例えば「平成 年 月 日論文内容とそれに関連した口頭試問を行った結果合格と認めた。」)を付け加えること。

Webでの即日公開を希望しない場合は、以下に公開可能とする日付を記入すること。
要旨公開可能日： 年 月 日以降